Statistical Papers 43, 401–422 (2002)

Statistical Papers

Nonparametric monitoring of financial time series by jump-preserving control charts

Ansgar Steland

Fakultät für Mathematik, Ruhr-Universität Bochum, Mathematik III, NA 3/71, 44780 Bochum, Germany

Received: September 28, 1999; revised version: July 18, 2000

Since structural changes in a possibly transformed financial time series may contain important information for investors and analysts, we consider the following problem of sequential econometrics. For a given time series we aim at detecting the first change-point where a jump of size a occurs, i.e., the mean changes from, say, m_0 to m_0+a and returns to m_0 after a possibly short period s. To address this problem, we study a Shewhart-type control chart based on a sequential version of the sigma filter, which extends kernel smoothers by employing *stochastic* weights depending on the process history to detect jumps in the data more accurately than classical approaches. We study both theoretical properties and performance issues. Concerning the statistical properties, it is important to know whether the normed delay time of the considered control chart is bounded, at least asymptotically. Extending known results for linear statistics employing deterministic weighting schemes, we establish an upper bound which holds if the memory of the chart tends to infinity. The performance of the proposed control charts is studied by simulations. We confine ourselves to some special models which try to mimic important features of real time series. Our empirical results provide some evidence that jump-preserving weights are preferable under certain circumstances.
Keywords: Digital image processing, EWMA control chart, financial econometrics, nonparametric smoothing, structural changes.

1 Introduction

Modern financial markets produce huge sequential data streams containing important information for investors and financial analysts, which have to make their decisions in a sequential fashion. Since structural changes may have severe implications on the decisions, there is a strong need to be in a position to detect change-points, where structural changes start, as soon as possible. Recent work on this subject has benefitted from results obtained in statistical process control for industrial production processes. It provides both a reasonable statistical framework and a rich collection of statistical methods to approach the problem. The main tools studied there are control charts which aim at monitoring a sequential stream of observations. A signal is given if there is evidence that the process is no longer in a state of statistical control but should be considered as out-of-control, e.g., since a jump in the mean (level) of the process is present.

Noting that the problem to detect a jump in an univariate time series is a one-dimensional version of the problem to detect an edge in an image, the present article studies a control procedure which is based on a nonparametric estimator called sigma filter. The sigma filter is designed to reproduce jumps and edges, respectively, accurately and has been develвopped and studied in digital image processing. In contrast to smoothing methods, the sigma filter has the attractive feature to reproduce jumps of certain heights exactly at the timepoint of their occurrence, provided the error distribution is concentrated on a finite interval. Thus, in this article we study whether sequential econometrics can benefit from the sigma filter.

We will now introduce the general setup. Suppose we are observing a possibly non-stationary stochastic process, $\{\widetilde{Y}(t) : t \in \mathcal{T}\}$, in continuous time, $\mathcal{T} = [t_0, \infty)$, with $E(|\widetilde{Y}(t)|) < \infty$ for all $t \in \mathcal{T}$. Assume that $\widetilde{Y}(t)$ can be decomposed in a possibly non-homogenous drift $m(t)$ and a zero-mean innovation process $\{\widetilde{\epsilon}(t) : t \in \mathcal{T}\}$, i.e.,

$$\widetilde{Y}(t) = m(t) + \widetilde{\epsilon}(t), \qquad (t \in \mathcal{T}). \tag{1}$$

This means, structural changes are modeled in the mean function $m(t)$. In the sequel we shall use the notation $\widetilde{Y}(t), \widetilde{\epsilon}_t, \ldots$ for quantities of the continuous time framework and will denote by $Y_n, \epsilon_n, \ldots$ the corresponding discrete time quantities.

The concern of this paper is to study a jump-preserving kernel method which can serve to analyze and detect structural change-points where the process mean $m(t)$ changes from a slow-varying behavior to a fast-varying behavior. This estimator, the sigma filter, has been first applied for nonparametric control chart design by Rafajłowicz (1996). As shown by Pawlak and Rafajłowicz (1999), it appears as a special case of the more general vertical regression approach discussed there in detail. For methods based on classical kernel estimators we refer to Schmid and Steland (1999) and the references given there.

We will address two types of deviations from an in-control state which are motivated by financial applications. Many (financial) processes can be modeled by stationary processes plus a smooth drift function for long time periods, which are interrupted by short periods of rapid changes of the process, sometimes lasting only a few days. This type of irregularity can even take the form of a type I outlier (cf. Fox (1972)). For instance, squared daily returns of stocks may exhibit such structures which are often modeled parametrically by GARCH-type models. Further, it may happen that for a long period of time the process is stationary, followed by a period where a smooth drift is present. Sometimes this irregular behavior is introduced by a rapid and sometimes even jump-like increase or decrease. It is of considerable interest to detect both types of change-points as soon as possible, since they indicate structural changes which often can be interpreted in economic terms. For instance, the value of an option on a security can change dramatically, when the volatility of the option's underlying increases or decreases.

To make the terms *slow-varying* and *fast-varying* precise, we shall study two different models. First, we may consider peak-like deviations

$$m(t) = m_0 + \mathbf{1}(t_n \geq \widetilde{t_q})\widetilde{\delta}(t), \tag{2}$$

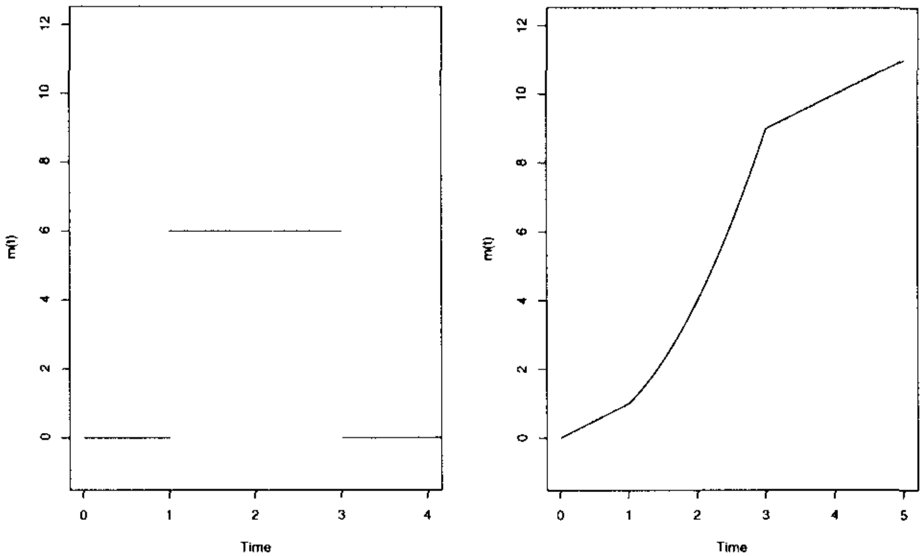
where $\widetilde{\delta}(t)$ is an arbitrary function with values in $\mathbb{R}$ and $\widetilde{\delta}(t) \neq 0$. Here and in the sequel $\mathbf{1}(A)$ stands for the indicator function for the expression A, i.e., $\mathbf{1}(A) = 1$ if A is true and $\mathbf{1}(A) = 0$ if A is false. In this model the change to a fast-varying behavior is modeled by a non-constant drift function. If $\widetilde{\delta}(t) = a \neq 0$ for all $t \in \mathcal{T}$, we obtain the classical change-point model. Second, we may model a fast change by imposing another assumption on $\widetilde{\delta}(t)$, namely

$$\widetilde{\delta}(t) \text{ is differentiable with } \widetilde{\delta}'(\widetilde{t_q}) \neq 0. \tag{3}$$

In this case $m(t)$ is differentiable with $m'(t) = 0$ for all $t \in [t_0, \widetilde{t_q})$ and $m'(\widetilde{t_q}) \neq 0$. In both models $\widetilde{t_q}$ is called change-point if $\widetilde{t_q} < \infty$. If there is some $s > 0$ so that $\widetilde{\delta}(t) = 0$ and $\widetilde{\delta}'(t) = 0$, respectively, for all $t > \widetilde{t_q} + s$, $\widetilde{t_q} + s$ may be called *second change-point*. Since our proposal to handle model (3) is to apply our proposal for model (2) after an appropriate transformation of the data, there will be no confusion about the assumptions on $\widetilde{\delta}(t)$. The two models are illustrated in figure 1.

The focus of this paper are sequential change-point detection procedures. However, there are various results on a posteriori methods where observations before, near, and after the change-point are available. Brodsky and Darkhovsky (1993) provide a review of various nonparametric methods, asymptotic results, and comparisons. A posteriori methods to detect jumps or sharp cusps nonparametrically have been studied by several authors. Ferger (1994a, 1994b, 1995) provides asymptotic results for change-point estimators based on U-statistics. Hall and Titterington (1992) proposed a method where left, right, and central smooths are calculated at each design point, where for the central smooth the nearest, say, m, data points are chosen, and for the left (right) smooth the corresponding nearest to the left (right) are selected. Based on these calculations a final estimate is proposed yielding an edge- and peak-preserving regression estimate. Methods based on wavelets have been considered in Ogden (1994) and Wang (1995).

Fig. 1 Illustration of the change-point models (2) and (3). The left panel shows a peak-like out-of-control situation ($s = 1$) with change points $t_q = 1$ and $t_q + s = 3$, whereas the right panel illustrates a mean function which is quadratic during the out-of-control period $[1, 3)$ and linear when the process is in control.



The organization of the paper is as follows. In Section 2 we provide a definition of our change-point model for the more realistic case that the process is sampled at discrete time points. In Section 3 the sigma filter and the associated detection procedure is introduced. Section 4 presents asymptotic results about our proposal. We show that under some regularity assumptions, in particular Cramér's condition, the sigma filter has exponential tails. Further, and more important, an upper bound for the normed delay time is established. Section 5 provides simulation studies to get some insight into the performance of the proposed method.

2 Change-point model in discrete time

We shall now assume that the process $\{\widetilde{Y}_t : t \in \mathcal{T}\}$ is observed at a sequence $\{t_1 < t_2 < \ldots\} \subset \mathcal{T}$ of fixed and ordered time points. To be consistent with model (1) assume

$$Y_n = m_n + \epsilon_n, \qquad n \in \mathbb{N},$$

where $Y_n = \widetilde{Y}(t_n)$, $m_n = m(t_n)$, $\epsilon_n = \widetilde{\epsilon}(t_n)$. We shall assume that $\{\widetilde{\epsilon}_n\}$ is a stationary sequence in discrete time $\mathbb{N}$. For a random design we may argue conditionally on $\{t_n\}$ if $\{Y_n\}$ and $\{t_n\}$ are independent. However, in the sequel we assume a fixed design with respect to time.

The discrete time analogon of model (2) is now given by

$$m_n = m_0 + \mathbf{1}(t_n \geq t_q)\delta_n, \qquad n \in \mathbb{N}, \tag{4}$$

where $\delta_n = \widetilde{\delta}(t_n) \neq 0$, $n \in \mathbb{N}$, and $t_q = \lfloor \widetilde{t_q} \rfloor + 1$. A simple discrete-time version of (3) is to consider the differential ratios

$$d_n = \frac{m_n - m_{n-1}}{t_n - t_{n-1}}, \qquad n \in \mathbb{N}.$$

Then $d_n = 0$ for all $t_n < t_q$, and it seems reasonable to define the change-point in this case by

$$d_{t_q} \neq 0.$$

Obviously, any control chart designed to detect a change-in-mean situation can be used to analyze model (4), simply by applying it to the data $\{(Y_n, t_n) : n \in \mathbb{N}\}$. Analogously, since the random variables

$$D_n = \frac{Y_n - Y_{n-1}}{t_n - t_{n-1}}, \qquad n \in \mathbb{N},$$

satisfy $E(D_n) = d_n$, we may simply employ such a control chart by applying it to the transformed data $\{(D_n, t_n) : n \geq 2\}$. To simplify notation we shall renumber this sequence and denote it again by $\{(D_n, t_n) : n \geq 1\}$.

3 A jump-preserving control chart

The application of Shewhart charts based on linear statistics $\sum_{i=1}^{n} w(t_i, t)Y_i$ evaluated at $t = t_n$, $\{w(t_i, t_n)\}$ being an appropriately defined weighting scheme, has one severe drawback. These estimators are primarily designed to provide *smooth* estimates for the process mean. If a fast, perhaps even jump-like change occurs, these estimators tend to oversmooth. Consequently, the corresponding control charts may perform badly. In digital image processing, where detection of edges, i.e., 'two-dimensional jumps', is an important task, the same problem arises. Lee (1983) proposed a solution, the sigma filter approach, whose one-dimensional analogon is given by

$$\widehat{m}_{R,n} = \widehat{m}_{R,n,h} = \frac{\sum_{i=1}^{n} \mathcal{K}_h(t_i - t_n) k_M(Y_i - Y_n) Y_i}{\sum_{j=1}^{n} \mathcal{K}_h(t_j - t_n) k_M(Y_j - Y_n)}. \tag{5}$$

Here, k and $\mathcal{K}$ are two non-negative kernels, and $k_M(\,\cdot\,)$ and $\mathcal{K}_h(\,\cdot\,)$ are the rescaled versions using positive bandwidths M and h, respectively. This estimator has been studied by Godtliebsen (1991) and Godtliebsen and Spjøtvoll (1991), and recently by Chiu et al. (1998) in a non-sequential setting.

The sigma filter smoothes the data not only with respect to time but also with respect to y. In particular, if both k and $\mathcal{K}$ are equal to the uniform kernel, $\widehat{m}_{R,n}$ puts a rectangle of sizes h and M, respectively, with upper right corner equal to (Y_n, t_n) on the data and estimates $\widehat{m}_{R,n}$ by the arithmetic average of all observations in the rectangle. If now a jump-like increase or decrease occurs, the windows moves, too. In this sense $\widehat{m}_{R,n}$ is jump-preserving.

Whereas the kernel $\mathcal{K}$ and the bandwidth h control the amount of past observations used to estimate the process mean, the kernel k and the bandwidth M are used to specify the estimator's sensitivity with respect to jumps. The approach studied here requires the statistician to fix some bandwidth M. The asymptotic results discussed below assume $h \to \infty$.

3.1 Construction of the control chart

Pawlak and Rafajłowicz (1999) proposed a control chart which gives a signal when the difference $\widehat{m}_{R,n+1} - \widehat{m}_{R,n}$ exceeds a threshold. In this paper we restrict attention to Shewhart-type control charts. For instance, the one-sided control chart is given by the stopping rule

$$N_h = \inf\{n \in \mathbb{N} : \widehat{m}_{R,n,h} > c\},$$

where c denotes an appropriate upper control limit. We compare the control statistic $\widehat{m}_{R,n,h}$ with a simple threshold, c, instead of a certain multiple of the in-control standard deviation, because the variance of the statistic $\widehat{m}_{R,n,h}$ is quite intractable. To detect change-points where the process $\{Y_n\}$ changes from a slow to a fast-varying behaviour as defined in Section 2, the procedure can be applied to the sequence $\{D_n\}$ instead of $\{Y_n\}$.

3.2 Choice of kernel.

Perhaps the most commonly used kernels for smoothing with respect to time are the Gaussian, the Epanechnikov, and the uniform kernel. It is worth noting that for the design of control charts the Laplace kernel,

$$\mathcal{K}_{Lap}(z) = \frac{1}{\sqrt{2}} \exp(-\sqrt{2}|z|)$$

is of some interest, too, because the associated kernel weights

$$w_h(t_i, t_n) = \mathcal{K}([t_i - t_n]/h) \;/\; \sum_{j=1}^{n} \mathcal{K}([t_j - t_n]/h) \tag{6}$$

satisfy

$$\lim_{n\to\infty} \frac{w_h(t_i, t_n)}{\lambda(1-\lambda)^{n-i}} = 1,$$

if the *equivalent bandwidth*

$$h_\lambda = -\sqrt{2}/\log(1-\lambda), \qquad \lambda \in (0,1] \tag{7}$$

is chosen and $t_\nu = \nu$ for all $\nu \in \mathbb{N}$ [Schmid & Steland (1999)]. In this sense, the corresponding linear statistic $\sum_{i=1}^n w(t_i, t_n) Y_i$ approximates the EWMA statistic, $\widehat{m}_{\mathrm{EWMA},n} = \sum_{i=1}^n \lambda(1-\lambda)^{n-i} Y_i$, which has been discussed in the literature to some extent. Using these choices the estimator $\widehat{m}_{R,n}$ is given by

$$\widehat{m}_{R,n} = \frac{\sum_{i=1}^n \lambda(1-\lambda)^{n-i} k_M(Y_i - Y_n) Y_i}{\sum_{i=1}^n \lambda(1-\lambda)^{n-i} k_M(Y_i - Y_n)}. \tag{8}$$

For k a natural choice may be the uniform kernel given by $k_U(z) = (1/2)\mathbf{1}(|z| \leq 1)$, or the *modified uniform kernel* defined as

$$k_\delta(z) = (1/2)\,\mathbf{1}(|z| \leq 1) + \delta, \tag{9}$$

$\delta > 0$, which attains a minimal weight $\delta > 0$ to each pair (Y_i, t_i). This kernel satisfies the assumptions of our main results and has been studied in our simulation study to some extent.

Remark 1 : Note that the sigma filter, $\widehat{m}_{R,n}$, converges to the classical kernel estimator $\sum_i w_h(t_i, t_n) Y_i$, employing kernel weights (6), if M tends to ∞. Thus, the parameter M controls the sensitivity with respect to jumps relative to the classical kernel estimator. In particular, the control statistic (8) can be regarded as a jump-preserving EWMA statistic, since

$$\lim_{M \to \infty, n \to \infty} \frac{\widehat{m}_{R,n}}{\widehat{m}_{\mathrm{EWMA},n}} = 1.$$

Remark 2 : Although in this paper we focus on kernel weights, any linear statistic with weights $w(t_i, t_n)$ can be made jump-preserving by using the weights

$$\widetilde{w}(t_i, t_n) = w(t_i, t_n) k_M(Y_i - Y_n) \;/\; \sum_{j=1}^n w(t_j, t_n) k_M(Y_j - Y_n), \quad (i = 1, \ldots, n).$$

Remark 3 : For some applications it might be attractive to replace the kernel weights $k_M(Y_i - Y_n)$ by some more general *(pseudo)-distance* $D(Y_i, Y_n; M)$. It is easy to modify the assumptions of our main results to ensure that they still hold.

4 Main results

Having in mind that the sigma filter relies on a highly nonlinear data adaptive method to set up a stochastic weighting scheme, the question arises whether the associated control chart possesses similar theoretical properties as, e.g, a Shewhart control chart based on a kernel smoother. Thus, this section provides our main results about the stochastic behavior of the proposed method. We start by estimating the tails of the sigma filter. Provided that the control limit is sufficiently large but fixed, the result says that the false-alarm probability approaches 0 at an exponential rate if h, the bandwidth determining the memory of the control chart, tends to ∞. Assuming the classical change-point model $m(t) = a\mathbf{1}(t \geq \widetilde{t_q})$, this result will be a building block to establish an upper bound, $\rho_0 = \rho_0(a)$, of the normed delay time which holds a.s. provided $h \to \infty$. The interpretation of these results is that for a Shewhart control chart based on the jump-preserving sigma filter with large h (long memory) the probability that the procedure stops after $[h\rho_0(a)] + 1$ tends to 0 at an exponential rate. Our results apply to independent but not necessarily identically distributed innovation processes.

4.1 Assumptions

The mathematical results of this article require some technical assumptions. Concerning the kernel functions k and $\mathcal{K}$, the following conditions are required.

(A1) The kernel $\mathcal{K}$ is a bounded density and symmetric around 0 with

$$\max_{z\in\mathbb{R}} \mathcal{K}(z) = \mathcal{K}(0) < \infty \qquad \text{and} \qquad \int \mathcal{K}(s)^2\, ds < \infty.$$

(A2) The kernel k is non-negative, bounded, integrable, and symmetric around 0 with
(i) $k(z) \geq k_{\min} > 0\ \forall z \in \mathbb{R}$.
(ii) $\max_{z\in\mathbb{R}} k(z) = k(0)$.

We assume that the innovation process forms a sequence of independent random variables. However, the results below do not require that they are identically distributed but allow for heteroscedasticity. We shall need *Cramér's condition.*

(A3) $\{Y_n\}$ satisfies Cramér's condition, i.e., there is a constant $c > 0$, such that

$$\sup_{\nu\in\mathbb{N}} E(\exp(c|Y_\nu|)) < \infty.$$

Assumption (A3) is strong. It implies existence of the moment generating function in a neighborhood of 0, therefore providing existence of all moments and exponential tail probabilities, i.e., $P(|Y_\nu| > x) = O(\exp(-ax))$, uniformly in ν, for all $x > 0$ and some constant $a > 0$ [cf. Petrov (1975), Lemma III.5]. If the observations Y_n are uniformly bounded, the proof given in the appendix can be simplified by applying Hoeffding's inequality. Finally, we confine ourselves to equidistant time designs.

(A4) Assume an equidistant design $t_\nu = \nu,\ \forall \nu \in \mathbb{N}$.

4.2 False-alarm probability when there is no drift

We will now provide the mathematical characterization of the in-control false-alarm behavior of the Shewhart chart based on $\{\widehat{m}_{R,n}\}$. Clearly, in-control means that $\{Y_n\}$ are independent zero-mean observations satisfying (A3), i.e., there is no underlying drift, $m(t) = 0$ for all $t \in \mathcal{T}$. We show that the estimator $\widehat{m}_{R,n}$ has exponential tails, asymptotically, a stochastic property which characterizes how fast the probability that $\widehat{m}_{R,n}$ has realizations outside an interval $(-\infty, x]$ approaches 0, if the bandwidth increases. We need this result to characterize the asymptotic out-of-control behavior of the normed delay in the next subsection, but it is of some interest in its own right, since this probability is the false-alarm probability, if x is a control limit.

The Theorem below also applies to the estimator

$$\widetilde{m}_{R,n} = \sum_{i=1}^{n} K_h(t_i - t_n) k_M(Y_i - Y_n) Y_i$$

employing weights only depending on Y_i and Y_n. Note that this estimator is also jump-preserving but has a simpler structure.

Theorem 1 *Let $\{Y_n\}$ be a sequence of independent random variables, and assume (A1)-(A4) and $a = 0$. If additionally,*

$$n/h \to \zeta > 0 \qquad \textit{as } n, h \to \infty,$$

then the following assertions hold true.

(i) There exists a constant $B > 0$ with

$$P(\widehat{m}_{R,n} > x) = O(\exp(-B \cdot h))$$

for every $x > 2M(R_k - 1) \sup_{\nu \in \mathbb{N}} E(Y_\nu^+)$, where $Y_\nu^+ = \max(0, Y_\nu)$.

(ii) There exists a constant $B > 0$ with

$$P(|\widetilde{m}_{R,n}| > x) = O(\exp(-B \cdot h)$$

for every

$$x > \mu_\Sigma = I(\zeta)[k(0) - k_{\min}] \sup_{\nu \in \mathbb{N}} E(Y_\nu^+)/M.$$

4.3 An upper bound for the normed delay

To establish an upper bound for the normed delay time which is not exceeded with probability 1 if $h \to \infty$, we will now assume the classical change-point model

$$Y_n = a \cdot \mathbf{1}(t_n \geq t_q) + \epsilon_n, \qquad n \in \mathbb{N}, \tag{10}$$

where without loss of generality the shift, a, is assumed to be positive, and $\{\epsilon, \epsilon_n\}$ is a sequence of i.i.d. zero-mean random variables with a common symmetric distribution function. ϵ stands for a generic copy and will be used to simplify the notation. The normed delay time is defined by

$$\rho_h = (N_h - t_q)/h$$

where $N_h = \inf\{n \in \mathbb{N} : |\widehat{m}_{R,n,h}| > c\}$ denotes the run length of the two-sided Shewhart control chart based on $\{\widehat{m}_{R,n,h}\}$ with upper control limit $c > 0$.

In the sequel we shall restrict ourself to the case that the average in-control run length, ARL_0, and hence the control limit c is fixed. For the case $ARL_0 \to \infty$ and related studies about asymptotic efficiencies we refer to Wu (1996) and the references given there. Brodsky and Darkhovsky (1993, Th. 4.2.8.) considered kernel charts with deterministic kernel weights providing a signal, if

$$|\sum_{i=n-h}^{n} \mathcal{K}_h(t_i - t_n)Y_i| > c$$

for some fixed threshold c. They proved that the associated normed delay time converges to ρ_0 with probability 1, if ρ_0 satisfies

$$\int_0^{\rho_0} \mathcal{K}(s)\, ds = \frac{c}{a},$$

provided $t_\nu = \nu \ \forall \nu \in \mathbb{N}$ Notice that this result implies that randomly stopped sums of i.i.d. random variables are asymptotically normal [Siegmund (1985), II.5.]

For the sigma filter the situation is more complicated, but it is still possible to provide an upper bound for the normed delay time for the special case $t_q = 1$. Then it is reasonable to redefine the normed delay as

$$\rho_h = N_h/h. \tag{11}$$

Assumption (A2) suggests the following definition. We shall call

$$R_k = \max_{z \in \mathbb{R}} k(z) \,/\, \min_{z \in \mathbb{R}} k(z) \tag{12}$$

the *kernel ratio of k*. Assumption (A2) ensures that R_k exists. For instance, the modified uniform kernel given by (9) provides a kernel ratio of $(\delta + 1/2)/\delta$.

Theorem 2 *Assume (A1)-(A4). Fix a threshold c. If (10) holds with $t_q = 1$, then a solution $\rho_0 = \rho_0(a)$ of*

$$\int_0^{\rho_0} \mathcal{K}(s)\, ds = \frac{cR_k}{2a - (R_k - 1)\sup_{\nu \in \mathbb{N}} E(\epsilon_\nu^+)} \tag{13}$$

provides an almost sure asymptotic upper bound for ρ_h as given in (11), in the sense that

$$P[\lim_{h \to \infty} \rho_h \le \rho_0] = 1,$$

since for any $\varepsilon > 0$ the probability that ρ_h exceeds ρ_0 by more than ε satisfies

$$P(\rho_h > \rho_0 + \varepsilon) = O(\exp\{-O(\varepsilon + O(1/h)) \cdot h\}),$$

and therefore converges to 0 at an exponential rate, as $h \to \infty$, i.e.,

Corollary 1 *The assertion of Theorem 2 remains true for the one-sided control chart where $\rho_h = N_h/h$, and $N_h = \inf\{n \in \mathbb{N} : \widehat{m}_{R,n,h} > c\}$.*

Remark 4 : Note that by definition of ρ_h the event $\{\rho_h > \rho_0\}$ equals $\{N_h > \rho_0 h\}$. Thus, the Theorem studies the tail behaviour of the stopping time N_h for $h \to \infty$.

5 Simulations

This section presents results about the performance of the methods discussed in this paper. Since many financial time series are affected by conditional heteroscedasticity, it may be not appropriate to restrict simulations to the case of Gaussian white noise. Thus, we also performed simulations assuming GARCH innovations. However, our primary goal was to compare the jump-preserving proposal to the widely used EWMA control chart. Our simulations provide some evidence that the jump-preserving proposal may be better than the EWMA, when the structural change present in the time series can be modeled by a rapid increase or decrease lasting for only a short period of time, at least if the change is not too small. For simplicity, in our simulations we modeled a structural change by two successive change-points, where the first change-point shifts the process mean to some level a, and after the second one the process is again a stationary zero-mean process.

More precisely, we generated time series according to the model

$$Y_n = a \cdot \mathbf{1}(t_q \leq t_n < t_q + s) + \epsilon_n, \qquad (n \geq -39),$$

with $t_\nu = \nu \; \forall \nu \in \mathbb{N}$, where $\{\epsilon_n\}$ is a sequence of zero-mean innovations. Here t_q is the first change-point in discrete time, and $t_q + s$ the second one. Whereas small values for s provide peaks, for $s \to \infty$ we obtain the classical change-point model. We studied the case $t_q = 1$, $s = 3$ for i.i.d. and GARCH innovations, respectively. Further, we analyzed how the choice of s affects the performance and investigated for the Gaussian case how the optimal choice of M depends on the noise level. To ensure a minimal amount of data for the control chart and to limit the dependence of simulated time series on starting values, a burn-in-period of size 40 was chosen.

To each generated time series we applied a Shewhart chart based on $\{\widehat{m}_{R,n}\}$ with $\widehat{m}_{R,n}$ as in (5). The kernel k was chosen as an uniform kernel and a modified uniform kernel, respectively, whereas we selected the Laplace kernel for K. As shown above, for large n and $M \to \infty$ the corresponding estimator is asymptotically equivalent to the EWMA control chart, and can be regarded as a jump-preserving EWMA control chart if M is fixed. The bandwidth M controlling the amount of smoothing with respect to the y-axis was chosen as $M \in \mathcal{M} = \{.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 5, 1000\}$, $M = 1000$ providing an approximation to the EWMA chart. To make results comparable, for h we used equivalent bandwidths according to (7) with $\lambda \in \{.02, .04, .06, .08, .1, .2\}$. This translates into bandwidths h ranging from 6.34 to 70.0, or, in terms of analyzing financial data on a daily basis, to time windows ranging from one week to three months.

For each parameter setting the upper control limit c yielding an in-control average run length equal to ξ was determined by a simulation with 500,000 repetitions. Taking account of the burn-in-period, the run length was re-defined as

$$N_h = \inf\{n \in \mathbb{N} : \widehat{m}_{H,n+1,h} > c(\xi)\}.$$

Since for some applications in finance small values of ξ seem to be appropriate, we used $\xi = 20$ corresponding to four business weeks. To take account of the control charts' long memory all simulations started with a pre-run of 40 observations. Each out-of-control average run length, defined here as the expected run length under any specification of the simulation model with $a \neq 0$, was estimated by a further simulation with at least 100,000 repetitions.

5.1 *Gaussian i.i.d. data.*

First, we generated samples with Gaussian i.i.d. innovations. The results are shown in table 1. We found that for small jump heights moderate to large values of M provide better results. For moderate to large jump heights smaller values of M provided better results. Taking account of imprecision of simulation results, it seems that this pattern applies for all λ-values considered here. Further, for fixed smoothing parameter λ and jump height a the average delay considered as a function of M seems to be U-shaped for moderate to large values of a. The effect of M on the average delay seems to be stronger than the effect of λ.

5.2 *Dependence of optimal M on the noise level for Gaussian data.*

The question arises how to choose the bandwidth M determining the estimator's sensitivity with respect to jumps. We expect that for a fixed peak length s and fixed bandwidth h the optimal M depends on the standard deviation of the innovation process. To get some insight into this relationship for $\sigma \in \{0.1, 0.2, \ldots, 1.0\}$ the optimal bandwith $M \in \mathcal{M}$ was determined via simulation, again for an in-control ARL equal to 20. The result for $s = 3$ and $\lambda = 0.06$ is given in figure 2.

Table 1 *Out-of-control average run lengths for Gaussian i.i.d. innovations. First change-point $t_q = 1$, second change-point $t_q + s = 3$. Optimal values of M printed in boldface.*

λ	a	M									
		0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	5.0	1000
0.02	0.5	16.07	15.97	15.72	15.48	15.14	14.75	14.49	14.48	**14.46**	14.60
0.02	1	10.58	10.35	9.98	9.65	9.29	**9.04**	9.05	9.30	9.62	9.96
0.02	2.5	2.46	2.33	2.23	2.17	2.11	**2.10**	2.22	2.51	3.32	3.97
0.04	0.5	16.17	16.22	15.74	15.39	15.04	14.59	14.41	14.39	14.95	**14.34**
0.04	1	10.66	10.49	9.98	9.48	9.07	8.79	**8.76**	8.88	9.67	9.45
0.04	2.5	2.47	2.37	2.20	2.11	2.07	**2.04**	2.11	2.35	3.07	3.30
0.06	0.5	16.16	16.09	16.14	15.72	14.99	14.61	14.41	14.30	14.33	**14.27**
0.06	1	10.67	10.42	10.30	9.75	9.03	8.70	**8.62**	8.73	9.02	9.08
0.06	2.5	2.50	2.37	2.26	2.14	2.03	**2.01**	2.07	2.25	2.68	2.90
0.08	0.5	16.29	16.60	15.81	15.67	15.06	14.71	14.35	**14.17**	14.45	14.21
0.08	1	10.78	10.90	10.11	9.70	9.05	8.71	**8.46**	8.57	8.91	8.87
0.08	2.5	2.52	2.48	2.23	2.13	2.02	**1.99**	2.04	2.17	2.52	2.67
0.10	0.5	17.45	16.20	16.10	15.43	15.03	16.48	14.35	14.25	**14.18**	14.19
0.10	1	11.69	10.59	10.30	9.56	9.02	9.90	**8.41**	8.43	8.55	8.66
0.10	2.5	2.74	2.43	2.27	2.12	**1.99**	2.17	2.00	2.12	2.34	2.47
0.20	0.5	16.25	16.24	16.18	15.78	15.22	14.81	14.44	14.35	**14.30**	14.38
0.20	1	10.78	10.91	10.61	9.95	9.15	8.69	8.37	**8.28**	8.29	8.37
0.20	2.5	2.53	2.58	2.37	2.16	1.99	1.91	**1.91**	1.94	2.02	2.08

5.3 GARCH Innovations.

Second, we modeled the innovations by a GARCH(1,1) model, i.e.,

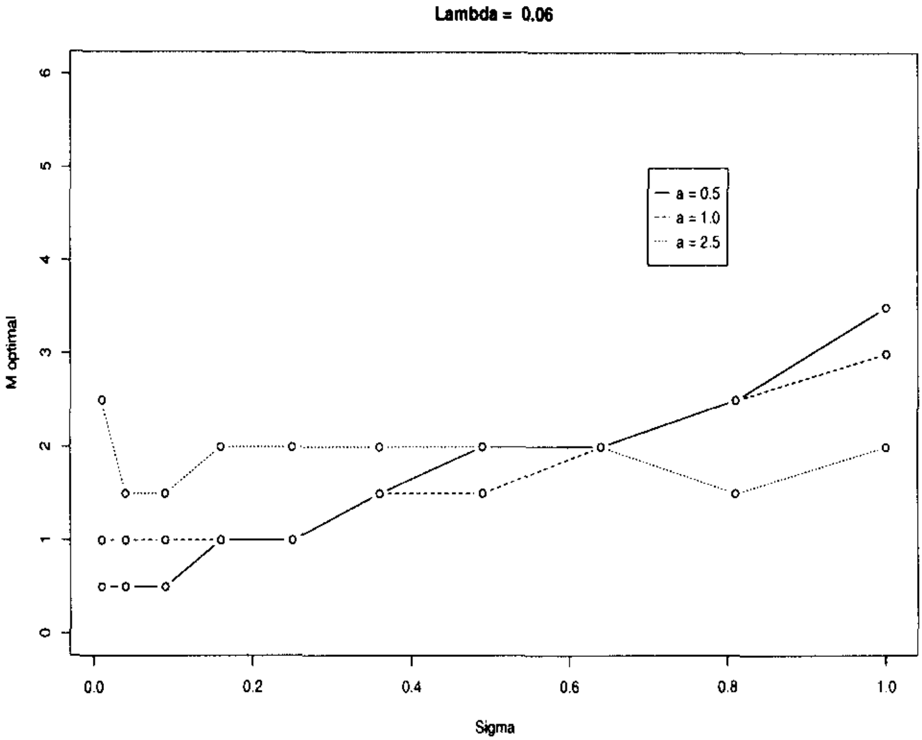
$$\epsilon_n = h_n \eta_n, \quad \text{with} \quad h_n^2 = \alpha_0 + \alpha_1 h_{n-1}^2 + \beta_1 \epsilon_{n-1}^2$$

for $n \geq 2$ and $h_1^2 = \alpha_0/(1 - \alpha_1 - \beta_1)$, where $\eta_n \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, and parameters α_0, α_1, and β_1. Empirical results suggest that $\alpha_0 = 0.1$, $\alpha_1 = 0.85$, $\beta_1 = 0.1$ is a realistic choice for financial time series. The results are shown in table 2. Again it seems that smaller values of M are preferable for detection of large jumps.

5.4 Peak-like deviations.

Further, we analyzed how the distance, s, between the first change-point and the second one affects the optimal choice of M. The corresponding results for Gaussian i.i.d. data and GARCH innovations are shown in table 3. Our results suggest that the optimal value for M may be increasing in s. Therefore, the jump-preserving proposal is particularly appealing when the statistician expects that there may be a second change-point right after the first one such that the process is in control again.

Fig. 2 Dependence of optimal M on the noise level. First change-point at $t = 1$, second one at $t = 4$. Optimal $M \in \mathcal{M}$ for $\sigma \in \{0.1, 0.2, \dots, 1.0\}$.



5.5 Results for the modified uniform kernel.

Since our theoretical results of Section 3 require that the kernel k used for smoothing with respect to the y-axis has an existing kernel ratio, we also considered modified uniform kernels k_δ as defined in (9) for $\delta \in \{1/9, 1/4, 1/2\}$ corresponding to kernel ratios of $R_k \in \{5\ 1/2, 3, 2\}$. Compared to the results for the uniform kernel it seems that there is now a slight tendency that smaller values of M are preferable and result in slightly smaller out-of-control average run lengths. However, the results are quite similar and do not suggest that the assumption of an existing kernel ratio is restrictive. Hence, we omit detailed simulation results.

6 Conclusions and Summary

Motivated by the fact that investors and financial analysts have to made their decisions sequentially based on the information contained in huge sequential data streams from capital markets, we studied whether sequential

Table 2 *Out-of-control average run lengths for GARCH innovations. Optimal values of M printed in boldface.*

λ	a	M									
		0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	5.0	1000
0.02	0.5	17.19	17.19	17.08	17.00	17.02	16.81	16.56	16.21	16.01	**15.93**
0.02	1	12.91	12.86	12.61	12.63	12.41	12.20	11.78	11.59	**11.57**	12.25
0.02	2.5	4.54	4.42	4.25	4.14	4.04	**3.99**	4.00	4.06	4.60	6.53
0.04	0.5	17.98	17.19	17.14	17.02	16.95	16.73	16.43	16.10	**15.72**	15.72
0.04	1	13.52	12.89	12.73	12.57	12.29	11.98	11.63	11.33	**11.21**	11.67
0.04	2.5	4.81	4.44	4.27	4.13	4.01	3.88	**3.79**	3.82	4.25	5.72
0.06	0.5	18.47	17.31	17.37	17.16	16.98	17.83	17.68	16.01	15.65	**15.56**
0.06	1	13.95	12.98	12.97	12.58	12.36	12.89	12.65	11.24	**11.06**	11.30
0.06	2.5	5.00	4.52	4.32	4.15	3.99	4.19	4.09	**3.73**	4.10	5.19
0.08	0.5	17.29	17.31	18.14	17.13	17.07	16.72	16.43	16.08	15.76	**15.59**
0.08	1	12.97	12.93	13.62	12.60	12.45	12.06	11.58	11.25	**10.97**	11.14
0.08	2.5	4.60	4.57	4.63	4.18	4.03	3.85	**3.70**	3.72	3.97	4.79
0.10	0.5	17.34	17.41	17.32	17.52	17.01	16.77	16.40	16.10	16.27	**15.56**
0.10	1	12.99	13.06	12.91	12.99	12.47	12.03	11.57	11.21	11.33	**10.92**
0.10	2.5	4.61	4.55	4.43	4.30	4.02	3.82	3.67	**3.66**	4.03	4.49
0.20	0.5	17.26	17.36	18.49	17.37	16.16	16.95	16.61	16.48	15.96	**15.71**
0.20	1	13.01	13.21	14.15	13.02	11.87	12.24	11.81	11.45	10.91	**10.76**
0.20	2.5	4.61	4.76	5.08	4.45	3.84	3.92	3.68	3.61	**3.56**	3.82

econometrics can benefit from the so-called sigma filter. The sigma filter aims at reproducing jumps in a data set more accurately than classical methods, e.g., kernel smoothers. Hence, it may be a useful tool to detect change points in financial data streams, which may indicate structural changes and can be important for financial decisions.

We establish two theoretical results characterizing the stochastic behavior of the detection procedure. We study both the tails of the sigma filter and the normed delay for the case that the bandwidth h, which controls the memory of the detection procedure, tends to infinity. Assuming the classical change-point model, it is shown that the normed delay is bounded by a deterministic constant with probability 1, if h tends to infinity.

Since sometimes structural changes are important but last only for a relatively short period of time, we performed computer simulations to study how a Shewhart control chart based on a sigma filter performs when confronted with the problem to detect peak-like structural changes. Our simulations suggest that the first occurrence of a change-point after a zero-mean period may be better detected by applying a control chart based on a sigma filter, when peak-like structural changes are present in the data. However, we also observed that in certain circumstances an approximation to the EWMA chart obtained by $M \to \infty$ performed equally well or even better.

Acknowledgments

The author is grateful to Prof. R. Pawlak, University of Manitoba, Canada, Prof. E. Rafajłowicz, Politechnika Wrocławska, Poland, and Prof. W. Schmid, European University Frankfurt (O), Germany, for stimulating discussions and helpful comments which improved the paper.

Table 3 *Out-of-control average run lengths for Gaussian i.i.d. data (top) and GARCH innovations (bottom). Optimization with respect to M over the grid $\mathcal{M}$. Optimal value for M in brackets, ∞ is short for $M = 1000$.*

λ	a	s = 1		2		4		8		64	
0.02	0.5	16.07	[4.00]	14.22	[4.00]	11.67	[5.00]	8.14	[5.00]	3.87	[5.00]
	1.0	11.53	[3.00]	8.93	[3.00]	5.72	[3.50]	2.95	[5.00]	2.12	[5.00]
	2.5	1.79	[2.50]	1.15	[3.50]	0.87	[4.00]	0.84	[4.00]	0.84	[4.00]
0.04	0.5	15.81	[∞]	14.39	[4.00]	11.59	[∞]	8.05	[∞]	4.37	[∞]
	1.0	*11.33*	[*3.00*]	*8.67*	[*3.50*]	*5.34*	[*4.00*]	*2.94*	[*4.00*]	*2.40*	[∞]
	2.5	1.75	[2.00]	1.19	[3.50]	0.97	[4.00]	0.96	[4.00]	0.96	[4.00]
0.06	0.5	15.91	[4.00]	14.26	[5.00]	11.44	[∞]	7.82	[∞]	4.70	[5.00]
	1.0	11.31	[3.00]	8.58	[3.50]	5.12	[4.00]	2.88	[5.00]	2.56	[5.00]
	2.5	1.74	[2.50]	1.19	[3.00]	1.05	[4.00]	1.04	[4.00]	1.04	[4.00]
0.08	0.5	15.91	[∞]	14.20	[4.00]	11.27	[∞]	7.80	[∞]	4.91	[∞]
	1.0	11.37	[3.50]	8.47	[3.50]	5.04	[4.00]	2.97	[∞]	2.67	[4.00]
	2.5	1.75	[2.50]	1.20	[3.50]	1.08	[3.50]	1.09	[3.50]	1.09	[3.50]
0.10	0.5	15.88	[4.00]	14.15	[5.00]	11.33	[∞]	7.75	[∞]	5.04	[∞]
	1.0	11.26	[3.50]	8.44	[4.00]	4.92	[5.00]	2.95	[5.00]	2.74	[5.00]
	2.5	1.73	[2.50]	1.23	[3.50]	1.12	[3.50]	1.12	[3.50]	1.11	[3.50]
0.20	0.5	16.00	[4.00]	14.34	[5.00]	11.50	[5.00]	8.03	[5.00]	5.45	[5.00]
	1.0	11.28	[3.50]	8.25	[5.00]	4.77	[5.00]	3.08	[5.00]	2.88	[5.00]
	2.5	1.74	[3.00]	1.23	[3.00]	1.17	[3.00]	1.16	[2.50]	1.17	[3.00]

λ	a	s = 1		2		4		8		64	
0.02	0.5	17.07	[∞]	15.97	[5.00]	13.82	[∞]	10.84	[∞]	4.94	[∞]
	1.0	13.80	[5.00]	11.49	[5.00]	8.33	[5.00]	4.78	[5.00]	2.81	[∞]
	2.5	3.53	[2.00]	2.08	[3.00]	1.35	[4.00]	1.15	[5.00]	1.16	[5.00]
0.04	0.5	17.00	[∞]	15.62	[∞]	13.45	[∞]	10.32	[∞]	5.47	[∞]
	1.0	13.54	[4.00]	11.14	[4.00]	7.77	[5.00]	4.59	[5.00]	3.09	[∞]
	2.5	3.45	[3.00]	2.05	[3.00]	1.33	[4.00]	1.26	[4.00]	1.26	[4.00]
0.06	0.5	16.78	[∞]	15.47	[∞]	13.28	[∞]	10.17	[∞]	5.79	[∞]
	1.0	13.48	[5.00]	11.00	[5.00]	7.55	[5.00]	4.49	[5.00]	3.26	[5.00]
	2.5	3.46	[2.50]	2.06	[2.50]	1.37	[4.00]	1.31	[4.00]	1.30	[4.00]
0.08	0.5	16.85	[∞]	15.57	[∞]	13.21	[∞]	10.07	[∞]	6.02	[∞]
	1.0	13.36	[5.00]	11.01	[5.00]	7.47	[5.00]	4.33	[5.00]	3.33	[5.00]
	2.5	3.47	[3.00]	1.98	[3.50]	1.38	[3.50]	1.33	[4.00]	1.33	[3.50]
0.10	0.5	16.99	[∞]	15.47	[∞]	13.12	[∞]	10.10	[∞]	6.13	[∞]
	1.0	13.41	[∞]	10.92	[∞]	7.43	[∞]	4.44	[∞]	3.45	[5.00]
	2.5	3.49	[3.50]	1.98	[3.50]	1.40	[3.50]	1.35	[3.50]	1.35	[3.50]
0.20	0.5	17.00	[∞]	15.60	[∞]	13.35	[∞]	10.35	[∞]	6.52	[∞]
	1.0	13.46	[∞]	10.78	[∞]	7.12	[∞]	4.43	[∞]	3.55	[5.00]
	2.5	3.41	[2.50]	1.96	[3.50]	1.43	[3.50]	1.39	[2.50]	1.38	[2.50]

A Proofs

Proof (of Theorem 1). By assumptions (A1) and (A4) $\{(t_n - t_i)/h : i = 1, \ldots, n\}$ forms an equidistant partition of $[0, (n-1)/h]$ with associated size

$1/h$. Let $b = \sup\{z \in \mathbb{R} : \mathcal{K}(z) > 0\}$. Then

$$\left|\sum_{i=1}^{n} \mathcal{K}_h(t_i - t_n) - \int_0^{(n-1)/h} \mathcal{K}(s)\, ds\right| \leq 1/(2h) \cdot \sup_{s \in \mathbb{R}} |\mathcal{K}'(s)| \min(b, (n-1)/h). \tag{14}$$

Hence, $\sum_i \mathcal{K}_h(t_i - t_n) = I(\zeta) + O(1/h)$ where $I(\zeta) = \int_0^\zeta \mathcal{K}(s)\, ds$. Similarly, under the same conditions,

$$\sum_{i=1}^{n} h^{-1}\mathcal{K}([t_i - t_n]/h)^2 = \int_0^\zeta \mathcal{K}^2(s)\, ds + O(1/h), \tag{15}$$

and, of course, for any $\alpha > 0$ we also have

$$\sum_{i=1}^{[\alpha h]} \mathcal{K}_h(t_i - t_{[\alpha h]}) = \int_0^\alpha \mathcal{K}(s)\, ds + O(1/h). \tag{16}$$

By (14) we have

$$\sum_{i=1}^{n} \mathcal{K}_h(t_i - t_n) k_M(Y_i - Y_n) \geq (k_{\min}/M)\{I(\zeta) + O(1/h)\}.$$

Therefore,

$$\begin{aligned}
&P(\widehat{m}_{R,n} > x) \\
&\quad = P\left[\frac{\sum_{i=1}^n \mathcal{K}_h(t_i - t_n) k_M(Y_i - Y_n) Y_i}{\sum_{i=1}^n \mathcal{K}_h(t_i - t_n) k_M(Y_i - Y_n)} > x\right] \\
&\quad \leq P\left[\sum_{i=1}^{n-1} \mathcal{K}_h(t_i - t_n) k_M(Y_i - Y_n) Y_i > \frac{x}{2}(k_{\min}/M)\{I(\zeta) + O(1/h)\}\right] \\
&\qquad + P[Y_n > (1/2)[\mathcal{K}(0)k(0)]^{-1} x k_{\min}\{I(\zeta) + O(1/h)\} \cdot h].
\end{aligned}$$

By Cramér's condition (A3) there are constants $b_1, b_2 > 0$ such that the second summand is bounded by $b_1 \exp(-b_2 h)$. Thus, it remains to provide a similar bound for the first term. Denote by $F_n(y)$ the distribution function of Y_n and observe that by independence of $Y_1, \ldots, Y_n$ the first term can be written as

$$\int P\left[\sum_{i=1}^{n-1} \mathcal{K}\left(\frac{t_i - t_n}{h}\right) k\left(\frac{Y_i - z}{M}\right) Y_i > \frac{x}{2} k_{\min}\left\{I(\zeta) + O(h^{-1})\right\} h\right] dF_n(z). \tag{17}$$

Obviously, the same argument yields

$$\begin{aligned}
P(\widetilde{m}_{R,n} > x) \leq \int P\left[\sum_{i=1}^{n-1} \mathcal{K}\left(\frac{t_i - t_n}{h}\right) k\left(\frac{Y_i - z}{M}\right) Y_i > xhM\right] dF_n(z) \\
+ b_1' \exp(-b_2' h)
\end{aligned} \tag{18}$$

for some positive constants b_1', b_2'. Thus, providing an exponential bound for the integrand of (17) will show (i) and (ii) provided the bound is uniform in $z \in \mathbb{R}$. We proceed as follows. Let

$$S_n(z) = \sum_{i=1}^{n-1} \mathcal{K}([t_i - t_n]/h)\xi_i(z)$$

where $\xi_i(z) = k([Y_i - z]/M)Y_i$. Note that all moments of the random variables $\xi_\nu(z)$ exist and are uniformly bounded in $z \in \mathbb{R}$ and $\nu \in \mathbb{N}$.

Assumption (A3), the fact that the kernel k is bounded, and (Petrov (1975), Lemma III.5) imply that for all $|t| \leq T$ and $g > (1/2)\ \sup_{\nu \in \mathbb{N}} \sup_z E([\xi_\nu(z) - E(\xi_\nu(z))]^2)$

$$E(\exp(t\mathcal{K}([t_i - t_n]/h)[\xi_i(z) - E(\xi_i(z))])) \leq \exp(\mathcal{K}([t_i - t_n]/h)^2 g t^2/2),$$

$i = 1, \ldots, n,\ n \in \mathbb{N}$. Markov's inequality provides for any $\eta > 0$

$$\begin{aligned} &P(S_n(z) - E(S_n(z)) > \eta) \\ &\leq \exp(-t\eta) \prod_{i=1}^{n-1} E(\exp(t\mathcal{K}([t_i - t_n]/h)[\xi_i(z) - E(\xi_i(z))]) \\ &\leq \exp(Kgt^2/2 - t\eta), \end{aligned}$$

where $K = \sum_{i=1}^{n-1} \mathcal{K}([t_i - t_n]/h)^2$. By minimizing the function $t \mapsto Kgt^2/2 - \eta t$ [cf. Brodsky & Darkhovsky (1993), p. 47.], we obtain

$$P(S_n(z) - E(S_n(z))| > \eta = \begin{cases} O(\exp(-\eta^2/(2gK))), & \eta \leq gTK \\ O(\exp(-\eta T/2)), & \eta > gTK \end{cases} \tag{19}$$

Note that $\{(t_n - t_i)/h = (n-i)/h : i = 1, \ldots, n-1\}$ provides an equidistant partition of $[0, (n-2)/h]$. Thus, the definition of K and (15) give $|K/h - \int_0^\zeta \mathcal{K}^2(s)\,ds| = O(1/h)$ which implies $|(K/h)/\int_0^\zeta \mathcal{K}^2(s)\,ds - 1| = O(1/h)$. Consequently, h/K is bounded away from 0 for sufficiently large h. To complete the proof we shall apply (19) with expressions for η satisfying $\eta \geq \eta' h$ for some $\eta' > 0$. In this case we obtain

$$\frac{\eta^2}{2gK} \geq \frac{(\eta')^2}{2g}\frac{h}{K}h \geq \text{const}\, h.$$

for a generic positive constant, uniformly in $z \in \mathbb{R}$. Consequently, there exists some constant $B > 0$ providing

$$P(S_n(z) - E(S_n(z)) > \eta) = O(\exp(-B \cdot h))$$

uniformly in $z \in \mathbb{R}$. Using the same arguments, we also obtain

$$P(-[S_n(z) - E(S_n(z))] > \eta) = O(\exp(-B' \cdot h))$$

where $B' > 0$. To prove assertion (ii), note that

$$[k_{\min} - k(0)] \sup_{\nu \in \mathbb{N}} E(Y_\nu^+) \leq E(\xi_i(z)) \leq [k(0) - k_{\min}] \sup_{\nu \in \mathbb{N}} E(Y_\nu^+),$$

since $E(Y_\nu^+) = -E(Y_\nu^-)$ by symmetry of Y_ν, $\nu \in \mathbb{N}$, and therefore

$$E(S_n(z)) \geq h[k_{\min} - k(0)] \sup_{\nu \in \mathbb{N}} E(Y_\nu^+)\{I(\zeta) + O(1/h)\} \tag{20}$$

$$E(S_n(z)) \leq h[k(0) - k_{\min}] \sup_{\nu \in \mathbb{N}} E(Y_\nu^+)\{I(\zeta) + O(1/h)\} \tag{21}$$

uniformly in $z \in \mathbb{R}$. Using (21) to bound the r.h.s. of (18), we obtain

$$\int P[S_n(z) - E(S_n(z)) > xhM - E(S_n(z))]\, dF_n(z)$$
$$\leq P[S_n(z) - E(S_n(z)) > (x - \mu_\Sigma + O(1/h)) \cdot Mh],$$

where $\mu_\Sigma = I(\zeta)[k(0) - k_{\min}] \sup_{\nu \in \mathbb{N}} E(Y_\nu^+)/M$. Applying (19) we obtain an exponential bound which is non-trivial, if $x > \mu_\Sigma$. Similarly, for positive constants b_1'', b_2''

$$P(-\widetilde{m}_n > x) \leq P(-[S_n(z) - E(S_n(z))] > xhM + E(S_n(z))) + b_1'' \exp(-b_2'' h),$$

where due to (20) the first term on the r.h.s. can be bounded by

$$P(-[S_n(z) - E(S_n(z))] > (x + \mu_\Sigma + O(1/h))Mh).$$

The resulting exponential bound is non-trivial if $x > 0$. Hence, $P(|\widetilde{m}_n| > x)$ is exponentially bounded in h as stated in (ii). To show (i), apply (19) to (17) with

$$\eta = x(1/2)(k_{\min}/M)\{I(\zeta) + O(1/h)\}h - E(S_n(z)).$$

Observe that $\eta = \eta' h$ with $\eta' > 0$, since the above estimate for $E(S_n(z))$ provides

$$\eta \geq [x(1/2)(k_{\min}/M) - [k(0) - k_{\min}] \sup_{\nu \in \mathbb{N}} E(Y_\nu^+)[I(\zeta) + O(1/h)] \cdot h.$$

The estimate is non-trivial for every $x > 2M(R_k - 1) \sup_{\nu \in \mathbb{N}} EY_\nu^+$, uniformly in $z \in \mathbb{R}$.

Proof (of Theorem 2). Let $\varepsilon > 0$. By definition of ρ_h and N_h we have

$$P(\rho_h - \rho_0 > \varepsilon) = P(N_h > (\rho_0 + \varepsilon)h) \leq P(|\widetilde{m}_{[(\rho_0+\varepsilon)h]}| \leq c)$$

where $[x]$ denotes the largest integer smaller than or equal to $x \in \mathbb{R}$. Note that

$$\sum_{i=1}^{[(\rho_0+\varepsilon)h]} K_h(t_i - t_{[(\rho_0+\varepsilon)h]}) k_M(Y_i - Y_{[(\rho_0+\varepsilon)h]}) \leq (k(0)/M)\,\{1/2 + O(1/h)\}.$$

Hence $P(\rho_h - \rho_0 > \varepsilon) \leq P(|\widetilde{m}_{[(\rho_0+\varepsilon)h]}| \leq u + O(1/h))$, where $u = ck(0)/(2M)$ and

$$\widetilde{m}_{[(\rho_0+\varepsilon)h]} = \sum_{i=1}^{[(\rho_0+\varepsilon)h]} \widetilde{w}_{[(\rho_0+\varepsilon)h],i} Y_i$$

with random weights

$$\widetilde{w}_{[(\rho_0+\varepsilon)h],i} = \mathcal{K}_h(t_i - t_{[(\rho_0+\varepsilon)h]})k_M(Y_i - Y_{[(\rho_0+\varepsilon)h]}) \qquad (i = 1, \dots, [(\rho_0+\varepsilon)h]).$$

Since $Y_i = a + \epsilon_i$ for $i = 1, \dots, [(\rho_0 + \varepsilon)h]$, an application of the inequality

$$|x + y| \leq z \ \Rightarrow \ |y| \geq |x| - z, \ x, y, z \in \mathbb{R},$$

provides the estimate

$$\begin{aligned}
&P(|\widetilde{m}_{[(\rho_0+\varepsilon)h]}| \leq u + O(1/h)) \\
&\leq P\left(\left|a \sum_{i=1}^{[(\rho_0+\varepsilon)h]} \widetilde{w}_{[(\rho_0+\varepsilon)h],i} + \sum_{i=1}^{[(\rho_0+\varepsilon)h]} \widetilde{w}_{[(\rho_0+\varepsilon)h],i}\epsilon_i\right| \leq u + O(1/h)\right) \\
&\leq P\left(\left|\sum_{i=1}^{[(\rho_0+\varepsilon)h]} \widetilde{w}_{[(\rho_0+\varepsilon)h],i}\epsilon_i\right| \geq a \ \times \right. \qquad (22) \\
&\left. \times \left\{ \sum_{i=1}^{[(\rho_0+\varepsilon)h]} \widetilde{w}_{[(\rho_0+\varepsilon)h],i} - \frac{u + \mu_\Sigma}{a} + O(1/h) \right\} + \mu_\Sigma \right)
\end{aligned}$$

where $\mu_\Sigma = I(\zeta)[k(0) - k_{\min}] \sup_{\nu \in \mathrm{N}} E\epsilon_\nu^+ / M$. Note that

$$\sum_{i=1}^{[(\rho_0+\varepsilon)h]} \widetilde{w}_{[(\rho_0+\varepsilon)h],i} \geq (k_{\min}/M) \sum_{i=1}^{[(\rho_0+\varepsilon)h]} \mathcal{K}_h(t_i - t_{[(\rho_0+\varepsilon)h]}).$$

Now, due to (16) we have

$$\sum_{i=1}^{[(\rho_0+\varepsilon)h]} \frac{1}{h}\mathcal{K}\left(\frac{[(\rho_0+\varepsilon)h] - i}{h}\right) = \int_0^{\rho_0+\varepsilon} \mathcal{K}(s)\,ds + O(1/h).$$

Recall the definitions of ρ_0, $R_k = k(0)/k_{\min}$, and $u = ck(0)/(2M)$ to obtain

$$\int_0^{\rho_0} \mathcal{K}(s)\,ds = \frac{cR_k + 2I(\zeta)[R_k - 1]\sup_{\nu \in \mathrm{N}} E\epsilon_\nu^+}{2a} + O(\varepsilon) + O(1/h).$$

Consequently,

$$\begin{aligned}
&P(\rho_h - \rho_0 > \varepsilon) \\
&\quad \leq P\left(\left|\sum_{i=1}^{[(\rho_0+\varepsilon)h]} \widetilde{w}_{[(\rho_0+\varepsilon)h],i}\epsilon_i\right| \geq \mu_\Sigma + a\{O(\varepsilon) + O(1/h)\}\right). \qquad (23)
\end{aligned}$$

Now an application of Theorem 1 (ii) provides

$$P(\rho_h - \rho_0 > \varepsilon) = O(\exp[-O(\varepsilon + O(1/h)) \cdot h]),$$

because the sum in (23) is given by

$$\sum_{i=1}^{[(\rho_0+\varepsilon)h]} \mathcal{K}([t_i - t_n]/h)k_M(\epsilon_i - \epsilon_{[(\rho_0+\varepsilon)h]})\epsilon_i,$$

which equals the estimator $\widetilde{m}_{[(\rho_0+\varepsilon)h]}$ applied to the sample $\epsilon_1, \ldots, \epsilon_{[(\rho_0+\varepsilon)h]}$. Therefore, we must apply Theorem 1 with $\zeta = \rho_0 + \varepsilon$. Since $I(\rho_0 + \varepsilon) = \int_0^{\rho_0} K(s)\, ds + O(\varepsilon)$, this finally gives

$$\int_0^{\rho_0} K(s)\, ds = \frac{cR_k}{2a - [R_k - 1] \sup_{\nu \in \mathbb{N}} E\epsilon_\nu^+}.$$

An application of the Theorem of Borel-Cantelli ensures that the event $\{\lim_{h\to\infty} \rho_h \le \rho_0\}$ has probability 1.

Proof (of Corollary 1). The proof goes along the lines of Theorem 2, since for the one-sided control chart the estimate (22) can be replaced by

$$P\left(a \sum_{i=1}^{[(\rho_0+\varepsilon)h]} \widetilde{w}_{[(\rho_0+\varepsilon)h],i} + \sum_{i=1}^{[(\rho_0+\varepsilon)h]} \widetilde{w}_{[(\rho_0+\varepsilon)h],i}\epsilon_i \le u\right)$$
$$\le P\left(\sum_{i=1}^{[(\rho_0+\varepsilon)h]} \widetilde{w}_{[(\rho_0+\varepsilon)h],i}(-\epsilon_i) \ge a\left\{\sum_{i=1}^{[(\rho_0+\varepsilon)h]} \widetilde{w}_{[(\rho_0+\varepsilon)h],i} - \frac{u+\mu_\Sigma}{a}\right\} + \mu_\Sigma\right).$$

Now the assertion follows, because $\{-\epsilon_n\}$ satisfies the assumptions of Theorem 1, and the weights depend on the underlying sequence only through the differences $\epsilon_i - \epsilon_{[(\rho_0+\varepsilon)h]}$, $i = 1, \ldots, [(\rho_0 + \varepsilon)h]$.

References

1. Brodsky, B. E. and Darkhovsky B. S. (1993), *Nonparametric Methods in Change-Point Problems*, Mathematics and its applications, **243**, Kluwer Academic Publishers, Dordrecht.
2. Chiu, C. K., Glad, I. K., Godtliebsen, F., and Marron, J. S. (1998), Edge-preserving smoothers for image processing, *Journal of the American Statistical Association*, **93**, 526-556.
3. Ferger, D. (1994a), On the power of nonparametric change-point tests, *Metrika*, **41**, 277-292.
4. Ferger, D. (1994b), Change-point estimators in case of small disorders, *Journal of Statistical Planing and Inference*, **40**, 33-49.
5. Ferger, D. (1995), Nonparametric tests for nonstandard change-point problems, *Annals of Statistics*, **23**, 1848-1861.
6. Fox, A. J. (1972), Outliers in time series, *Journal of the Royal Statistical Society, Series B*, **43**, 350-363.
7. Godtliebsen, F. (1991), Noise reduction using Markov random fields, *Journal of Magnetic Resonance*, **92**, 102-114.
8. Godtliebsen, F. and Spjøtvoll, E. (1991), Comparison of statistical methods in MR imaging, *International Journal of Imaging Systems and Technology*, **3**, 33-39.
9. Hall, P. and Titterington, D. M. (1992), Edge-preserving and peak-preserving smoothing, *Journal of the American Statistical Association*, **34**, 4, 429-440.
10. Härdle, W. (1990), *Applied Nonparametric Regression*, Cambridge University Press, Cambridge, UK.

11. Kramer, H. G. and Schmid, W. (1998), On the average delay of control schemes, In *Risk Measurements, Econometrics and Neural Networks*, Physica, Contributions to Economics, 341-360.
12. Lee, J. S. (1983), Digital image smoothing and the sigma filter, *Computer Vision, Graphics and Image Processing*, **24**, 255-269.
13. Lombard, F. (1988), Detecting change-points by Fourier analysis, *Technometrics*, 30, 305-310.
14. Lucas, J. M. and Saccucci, M. S. (1990), Exponentially weighted moving average control schemes: properties and enhancements, *Technometrics*, **32**, 1-12.
15. Müller, H. G. (1992), Change-points in nonparametric regression analysis, *Annals of Statistics*, **20**, 737-761.
16. Muth, J. F. (1960), Optimal properties of exponentially weighted forecasts, *Journal of the American Statistical Association*, **74**, 365-367.
17. Neumann, M. H. (1997), Optimal change-point estimation in inverse problems, *Scandinavian Journal of Statistics*, **24**, 503-521.
18. Ogden, R. T. (1994), *Wavelet thresholding in nonparametric regression with change-point applications*, Ph. D. thesis, Texas A&M University.
19. Page, E. S. (1954), Continuous inspection schemes, *Biometrika*, **41**, 100-115.
20. Pawlak, M. and Rafajłowicz, E. (1999), Vertically weighted regression - a tool for nonlinear data analysis and constructing control charts, *Working Paper.*
21. Petrov, V. V. (1975), *Sums of independent random variables*, Transl. from the Russ. by A. A. Brown, Springer, Berlin.
22. Rafajłowicz, E. (1996), Model free control charts for continuous time production processes, Politechnika Wrocławska, *Technical Report* No. 56/96.
23. Roberts, S. W. (1959), Control chart tests based on geometric moving averages, *Technometrics*, **1**, 239-250.
24. Schmid, W. (1997b), On EWMA charts for time series, In *Frontiers in Statistical Quality Control*, Lenz, H. J. and Wilrich, P. Th. (eds.), 115-137, Physica-Verlag, Heidelberg.
25. Schmid, W. (1997a), CUSUM control schemes for Gaussian processes, *Statistical Papers*, **38**, 191-217.
26. Schmid, W. and Schöne, A. (1997), Some properties of the EWMA control chart in the presence of autocorrelation, *Annals of Statistics*, **25**, 3, 1277-1283.
27. Schmid, W. and Steland, A. (1999), Sequential Control of Non-Stationary Processes by Nonparametric Kernel Control Charts, *Working Paper.*
28. Siegmund, D. (1985), *Sequential Analysis*, Springer, New York.
29. Wand, M.P. and Jones, M.C. (1990), *Kernel Smoothing*, Chapman and Hall, London.
30. Wang, Y. (1995), Jump and sharp cusp detection by wavelets, *Biometrika*, **82**, 385-397.
31. Wu, Y. (1996), A less sensitive linear detector for the change-point based on kernel smoothing method, *Metrika*, **43**, 43-55.
32. Wu, J. S. and Chu, C. K. (1993), Kernel-type estimators of jump points and values of a regression function, *Annals of Statistics*, **21**, 3, 1545-1566.